Web-Based Access and Visualization of Hazardous Air Pollutants

Jürgen Symanzik (1),* David Wong (2), Jingfang Wang (2), Daniel B Carr (2), Tracey J Woodruff (3), Daniel A Axelrad (3)
(1) Department of Mathematics and Statistics, Utah State University, Logan, UT (2) George Mason University, Fairfax, VA; (3) Office of Policy, US Environmental Protection Agency, Washington, DC

Abstract

This paper reports on a project to provide Web-based access to the US Environmental Protection Agency's (EPA's) extensive model-based summaries of hazardous air pollutants (HAPs). As part of EPA's Cumulative Exposure Project, long-term cumulative concentrations of 148 HAPs for the 60,803 census tracts in the 48 contiguous states have been modeled for 1990. The model results include estimates and confidence bounds that assess the estimated uncertainties for each of the HAPs in each census tract. The project challenge was to concisely display 148×60,803 (8,998,844) estimates along with uncertainty bounds. The project goal was to make these statistical summaries accessible to the public as statistical tables and graphs. The Web provides an easy way to make this information electronically accessible. One big challenge is to make the summaries conceptually accessible. The most difficult part of this is to communicate an understanding of the underlying data limitations, the modeling process, and how to interpret the model results. The easier part of ensuring conceptual accessibility is facilitating navigation through the summaries and consideration of values within a large context. Our approach allows the user to select a HAP and "drill down" through the levels of a geopolitical hierarchy. The hierarchy consists of states within the United States, counties within states, and census tracts within counties. Our Web-based approach also attended to the design of tables and graphics with the intent to make them more readable and useful. For tables, our approach focused on perceptual details such as rounding and foreground-background contrast. For graphics, our approach provided spatial context through the use of recently developed templates called linked micromap plots. Both tables and micromaps provide a hierarchically clickable drill-down to finer details. This provides fast answers to questions about the air quality in any given region in the contiguous United States.

Keywords: Cumulative Exposure Project, HAPs, linked micromap plots, micromaps, Graphics Production Library

Introduction

Over the last few years, researchers have developed many improvements that make statistical graphics more accessible to the general public. These improvements include making statistical summaries more visual and providing more information at one time. Research in this area involved exploring the conversion of statistical tables into plots (1), new ways to display geographically referenced data (2), and, in particular, the

* Jürgen Symanzik, Utah State University, Dept. of Mathematics and Statistics, 3900 Old Main Hill, Logan, UT 84322-3900 USA; (p) 435-797-0696; (f) 435-797-1822; E-mail: symanzik@sunfs.math.usu.edu

development of linked micromap (LM) plots, often simply called micromaps (3–5). Another recent development is the Java-based Graphics Production Library (GPL) (Bureau of Labor Statistics, Washington, DC) for the Web-based distribution of interactive statistical graphics (6). The GPL can be used to distribute federal statistical summaries such as the description of hazardous air pollutants (HAPs).

Staff at the US Environmental Protection Agency (EPA) and contractors modeled 1990 long-term cumulative concentrations for 148 HAPs at the census tract level. Because there are 60,803 census tracts in the contiguous United States, this resulted in 148×60,803 (8,998,844) estimates, along with upper and lower confidence bounds for each estimate. The main goal of this project was to provide Web-based access to these HAP data. We focused on providing a concise display that offered easy access to the data. Another goal was to make the model results easily understandable to an audience not familiar with statistics. To achieve our goals, we developed interactive tables and extended the GPL by adding micromaps. The micromaps serve as a geographic navigational "drill-down" tool as well as being meaningful statistical overviews in their own right. The ability to drill down from the national overview showing states, to a state overview showing counties, and then to a county view showing census tracts, provides rapid access to the fine-grained detail in this substantial dataset.

In the next section of this paper we describe EPA's Cumulative Exposure Project (CEP). In the section entitled "Graphical Statistical Components," we describe components, i.e., GPL and micromaps, that were used to construct the CEP Web site. The section following that provides deeper insights into the CEP Web site and the user's point of view. We finish with a discussion of work that has been done to date and that which is still to come. Additional details and a different set of micromap displays and screenshots from the CEP Web site can be found in Symanzik et al. (7).

EPA's Cumulative Exposure Project

Much of the characterization of air pollution has focused on air pollutants designated as "criteria pollutants" in the Clean Air Act, such as particulate matter (PM), ozone, and lead (8). This is largely due to the obvious health effects demonstrated by major pollution episodes, such as those in Donora, Pennsylvania, and London, England (9,10), and the extensive availability of monitoring data to use in assessing health effects. Relatively little is known about the potential health effects of other air pollutants, a number of which are designated as HAPs in the Clean Air Act. HAPs have been associated (mostly through occupational and animal studies) with a variety of adverse health outcomes, including cancer effects and noncancer neurological, reproductive, and developmental effects (11).

Past analyses have relied on limited emissions and monitoring data and some modeling to assess public health impacts of air toxics. Some studies have attempted to assess differential impacts of air toxics on communities of color using emissions estimates, mostly from the EPA's Toxics Release Inventory (TRI), which contains emissions estimates from major manufacturers in the United States (12,13). Other analyses have attempted to characterize the potential public health impacts of air toxics (13–17). One set of studies evaluates potential noncancer health risk by using monitoring data and concentrations estimated by dispersion modeling of emissions from a subset of commercial and industrial facilities (14–16). These studies found that outdoor

concentrations were often greater than benchmarks representing thresholds for potential public health impacts. However, these studies were not comprehensive in their scope, due either to lack of monitoring data (as described in reference 18) or to lack of emissions data.

A recent analysis by Woodruff et al. (19), as part of the EPA's CEP, has assessed the potential public health implications of air toxics across the United States for 1990. The analysis by Woodruff et al. uses modeled outdoor concentrations of air toxics across the contiguous United States (20) to help compensate for the lack of monitoring data on outdoor concentrations. Emissions data from stationary and mobile sources are used as inputs into a dispersion model that estimates 1990 average outdoor concentrations of 148 air toxics for every census tract in the contiguous United States. The estimated outdoor concentrations from the analysis are used as a reasonable proxy for potential exposure when making relative comparisons of hazard and performing screening-level analysis. The analysis by Woodruff et al. found that many estimated concentrations are above previously defined benchmark concentrations representing thresholds of concern for potential adverse public health impacts (19,21).

Estimating 1990 Outdoor Concentrations of Hazardous Air Pollutants

Outdoor concentrations of HAPs were estimated using a Gaussian dispersion model (20,22). This model—the Assessment System for Population Exposure Nationwide (ASPEN)—is a modified version of EPA's Human Exposure Model (22), a standard tool designed to model long-term concentrations over large spatial scales. Long-term average concentrations of HAPs were calculated at the census tract level[1] based on emissions rates of the HAPs and frequencies of various meteorological conditions, including wind speed, wind direction, and atmospheric stability. In addition, the model used in this analysis incorporates simplified treatment of atmospheric processes such as decay, secondary formation, and deposition.

The choice of pollutants for modeling was based on the list of 189 HAPs in section 112 of the 1990 Clean Air Act Amendments. A baseline year of 1990 was selected for modeling. Available emissions data were reviewed and appropriate data were identified for 148 HAPs.

A national inventory of HAP emissions was developed for this study as a required input to the dispersion model. For large manufacturing sources, emissions data contained in EPA's TRI were used (23). Emissions estimates were developed for other sources, such as large combustion sources, automobiles, and dry cleaners, using EPA's extensive national inventories of 1990 emissions of total volatile organic compounds (VOCs) and PM (24,25). HAP emissions were derived from VOC and PM emissions estimates by applying industry-specific and process-specific estimates of the presence of particular HAPs in particular VOC or PM emissions streams (20). Alaska and Hawaii are not included in this study because the national VOC and PM emissions inventories do not include data for these states.

The dispersion model accounted for long-term concentrations of HAPs attributable to current (i.e., 1990) anthropogenic emissions within 50 kilometers of each census tract centroid. For 28 HAPs, estimated outdoor concentrations also included a "background"

[1] The 60,803 census tracts in the contiguous United States vary in physical size but typically have approximately 4,000 residents.

component attributable to long-range transport, re-suspension of historical emissions, and natural sources derived from measurements taken at "clean air" locations remote from the impact of local anthropogenic sources (20). Compared to available air toxics monitoring data for 1990, 1990 modeled concentrations are typically of the correct magnitude, with a general tendency to underestimate the measured ambient concentrations.

EPA's CEP Web site (http://www.epa.gov/CumulativeExposure) was designed specifically to provide further insight into statistical methods and methodologies and answer questions related to air toxics. In addition to providing explanatory texts, documents, and external links that relate to the material described in this section, one of the main goals for the CEP Web site was to provide information about the estimated air toxics data. An essential part of assessing the data is the option to evaluate them visually. The remainder of this paper describes the work done to present the data and the development of the CEP Web site.

Graphical Statistical Components

This section introduces the main graphical statistical components that are used for the Web-based access and visualization of HAPs through EPA's CEP Web site.

The Graphics Production Library

The GPL is a Java class library of graphics routines that make it possible (and convenient) to create and modify statistical graphics on the Web (see http://www.monumental.com/dan_rope/gpl/) (6). The GPL was initially developed within the Bureau of Labor Statistics (BLS) to facilitate the Web-based distribution of the Bureau's statistical summaries. It has interactive features such as dragging and dropping data columns onto each other to allow easier comparisons of the data, reordering and rescaling of panels, and panning and zooming. Thus, it considerably extends the static features but otherwise closely follows the row-labeled plots of Carr (1). Recent recommendations on statistical graphics, as given in Cleveland (26,27) have also been followed during the design of the GPL. Moreover, the GPL makes it possible to add metadata—i.e., add links to articles associated with the data or include warning flags within a data display. The GPL currently supports three types of graphics displays: bar plots, dot plots, and time series graphics. Other types of graphics have been planned but have not been implemented so far by BLS. Unfortunately, the GPL cannot be used to draw maps and link statistical data to them; however, this is one of the features that have been planned.

Micromaps

Linked micromap (LM) plots, often simply called micromaps, provide a new statistical paradigm for the viewing of spatially referenced statistical summaries in their spatial context. Full details on LM plots can be found in Carr et al. (3–5). Using LM plots for the 50 states within the United States provides an alternative to displaying all statistical information on a single choropleth map. Instead, several small maps (ten maps in our 50-state example) are drawn. The associated statistical data are arranged according to a particular criterion (e.g., from highest to lowest or in alphabetical order by corresponding geographical region). Next, the five highest values of the data are drawn in a statistical plot (e.g., dot plot, bar plot, box plot, plot with confidence bounds, time series plot) on the right side of the first small map. For each data point, a different color

is used in the statistical plot. The corresponding regions (in this case five states) are highlighted in the same colors on the first small map. All other states remain uncolored in this map. The same is done for the next five highest remaining data points in the second small map and associated statistical display. This process continues until all data points/regions have been plotted/highlighted.

This splitting into several maps makes obvious the locations of the high, middle, or low observations. It becomes possible to judge if there are any geographic clusters or if the underlying measurements are randomly spread over the area under consideration. LM plots can display multiple statistical variables at a time. Examples of micromaps and S-PLUS (MathSoft, Seattle, WA) code written to create them can be found at ftp://galaxy.gmu.edu/pub/dcarr/newsletter/micromap/.

Figure 1 shows a sample LM plot created using S-PLUS. This county-level micromap of Pennsylvania generally follows the design principle described above. However, we had to find an (almost) symmetric display for 67 counties. We ended up with 16 maps, 13 of them with four counties and 3 of them with five counties. Moreover, the layout has been split into four quarters. In addition to simply highlighting the individual four or five counties from the associated statistical display in each map, all 16 or 17 counties that fall into the corresponding quarter are highlighted on this map. This makes it easier to understand the spatial structure. For example, the viewer can immediately grasp that the highest benzene concentrations have been modeled for counties surrounding the major cities Philadelphia (e.g., Philadelphia County, Delaware, Montgomery, and Bucks), Harrisburg (e.g., Dauphin and Cumberland), and Pittsburgh (e.g., Allegheny), while the lowest benzene concentrations have been modeled for counties in the sparsely populated Pennsylvania-New York border region (northern border of map).

The CEP Web Site

The User's Point of View

In addition to providing access to explanatory texts and entire CEP-related documents, the main purpose of the CEP Web site is to provide fast and easy access to data on the 148 HAPs at different spatial resolutions ranging from the US level (top) to the census tract level (bottom).

Three mechanisms have been designed for selecting main features of the CEP Web site and for maneuvering from the US down to the census tract level and up again. Standard menus in the upper-right part of the Web page allow the user to select the representation (data tables, micromaps, or raw data), one of the 148 HAPs, a state and, based on this selection, a county for which the data should be displayed (see Figure 2). This navigation and selection menu remains permanently visible, independently from the current statistical display.

The second tool that has been implemented to drill down through the levels of a geopolitical hierarchy makes use of interactive tables that display statistical data and serve as navigational tools at the same time. Figure 2 shows such a table at the county level for Pennsylvania. The user can mouse-click on any of the listed counties and a new table will appear, displaying data (including a 90% confidence interval) for all census tracts within the selected county. Small arrows pointing upward and downward allow

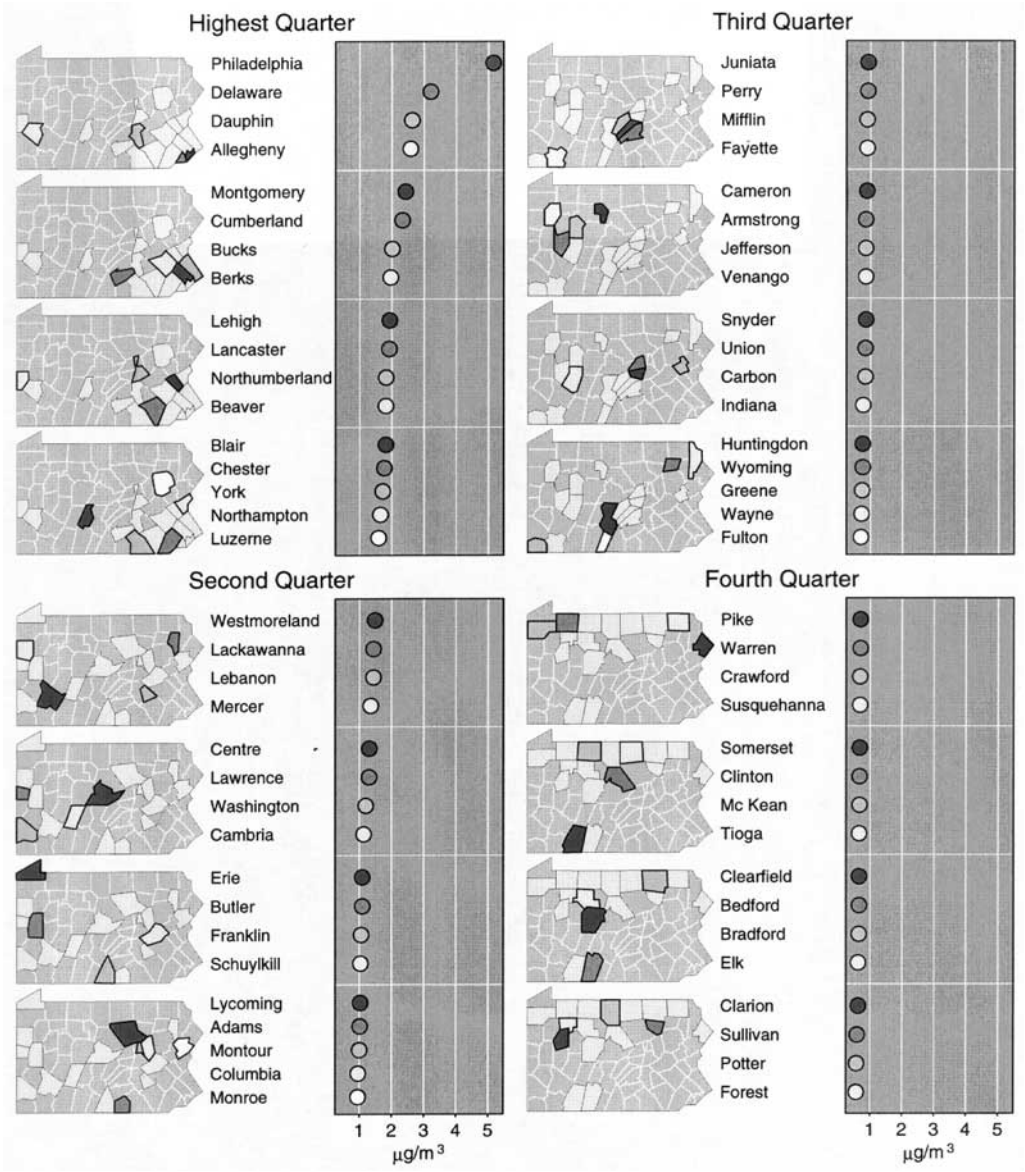
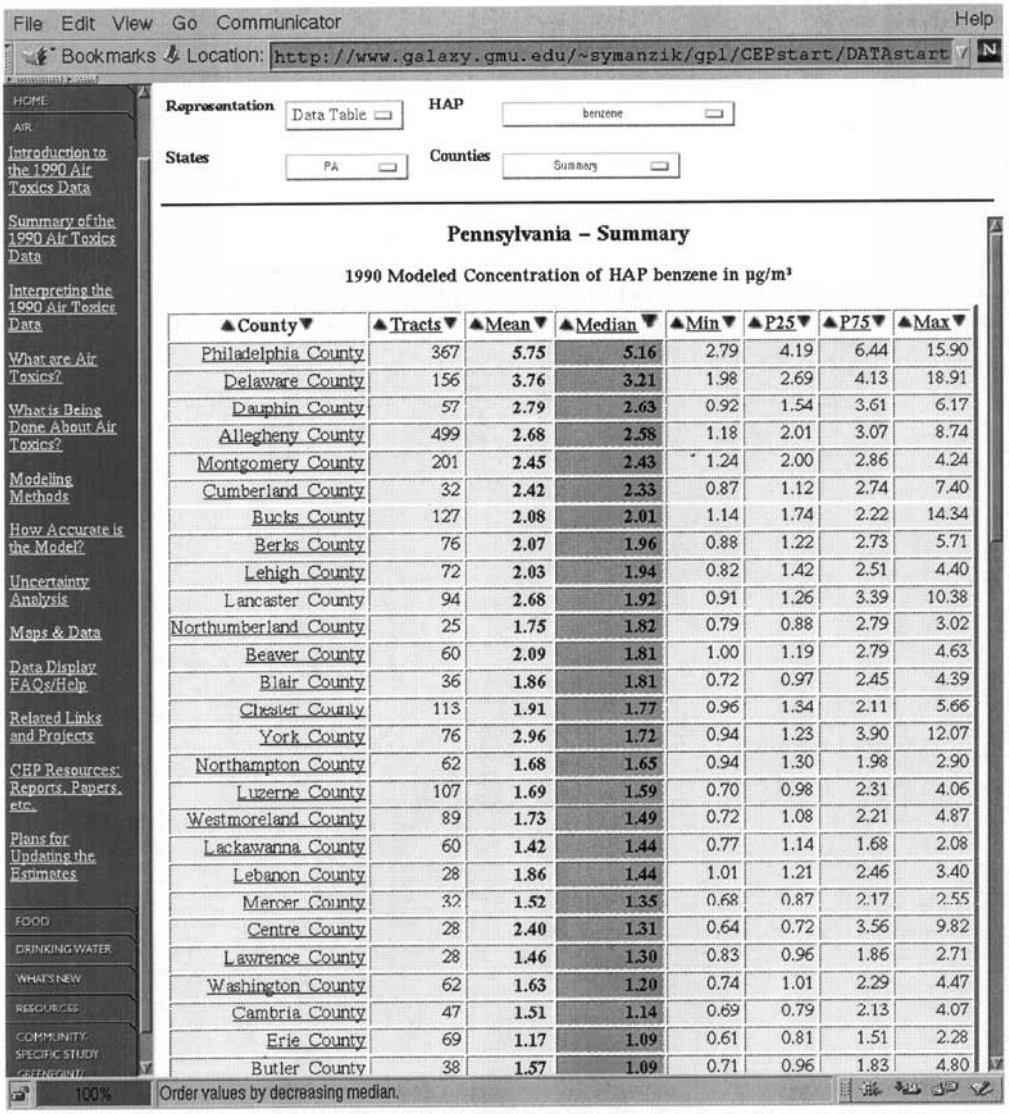


Figure 1 Micromap display at the state level for Pennsylvania, showing all of its 67 counties. Displayed is the median (with respect to the census tract estimates within each county) of the modeled 1990 benzene concentration in micrograms per cubic meter. Counties are ordered from highest to lowest median benzene concentration.

the user to rearrange the rows of the table in increasing or decreasing order as defined by the selected criterion. In the current view, the data are ordered from highest to lowest median benzene concentration. This is indicated through the larger downward arrow and the different-colored background of this data column. Through this interactivity, valuable information can be found in a larger table within seconds instead of



County	Tracts	Mean	Median	Min	P25	P75	Max
Philadelphia County	367	5.75	5.16	2.79	4.19	6.44	15.90
Delaware County	156	3.76	3.21	1.98	2.69	4.13	18.91
Dauphin County	57	2.79	2.63	0.92	1.54	3.61	6.17
Allegheny County	499	2.68	2.58	1.18	2.01	3.07	8.74
Montgomery County	201	2.45	2.43	1.24	2.00	2.86	4.24
Cumberland County	32	2.42	2.33	0.87	1.12	2.74	7.40
Bucks County	127	2.08	2.01	1.14	1.74	2.22	14.34
Berks County	76	2.07	1.96	0.88	1.22	2.73	5.71
Lehigh County	72	2.03	1.94	0.82	1.42	2.51	4.40
Lancaster County	94	2.68	1.92	0.91	1.26	3.39	10.38
Northumberland County	25	1.75	1.82	0.79	0.88	2.79	3.02
Beaver County	60	2.09	1.81	1.00	1.19	2.79	4.63
Blair County	36	1.86	1.81	0.72	0.97	2.45	4.39
Chester County	113	1.91	1.77	0.96	1.34	2.11	5.66
York County	76	2.96	1.72	0.94	1.23	3.90	12.07
Northampton County	62	1.68	1.65	0.94	1.30	1.98	2.90
Luzerne County	107	1.69	1.59	0.70	0.98	2.31	4.06
Westmoreland County	89	1.73	1.49	0.72	1.08	2.21	4.87
Lackawanna County	60	1.42	1.44	0.77	1.14	1.68	2.08
Lebanon County	28	1.86	1.44	1.01	1.21	2.46	3.40
Mercer County	32	1.52	1.35	0.68	0.87	2.17	2.55
Centre County	28	2.40	1.31	0.64	0.72	3.56	9.82
Lawrence County	28	1.46	1.30	0.83	0.96	1.86	2.71
Washington County	62	1.63	1.20	0.74	1.01	2.29	4.47
Cambria County	47	1.51	1.14	0.69	0.79	2.13	4.07
Erie County	69	1.17	1.09	0.61	0.81	1.51	2.28
Butler County	38	1.57	1.09	0.71	0.96	1.83	4.80

Figure 2 Tabular display at the state level for Pennsylvania showing the modeled 1990 benzene concentration in micrograms per cubic meter. Displayed are the number of census tracts and summary statistics (e.g., mean and median) with respect to the census tract estimates within each county. Counties are ordered from highest to lowest median benzene concentration. While only 27 of Pennsylvania's 67 counties are visible in this figure, the remaining 40 counties are accessible through the right scrollbar when looking at these data on the Web.

having to search sequentially through each table column to find largest, smallest, and—even more complicated—center values.

It should be noted that, when designing our tabular display, we have paid particular attention to recommendations from the cognitive sciences. Numbers have been rounded to two significant digits (with respect to the smallest number in any given

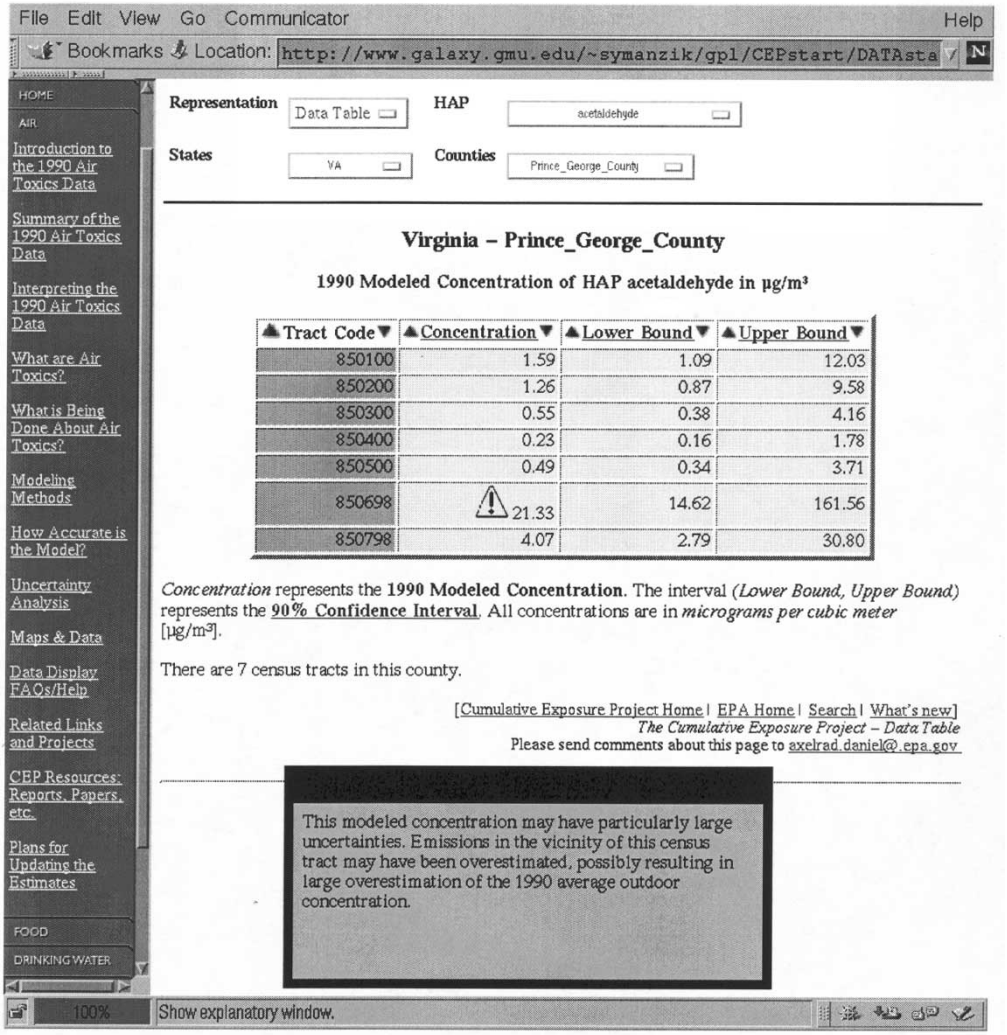


Figure 3 Tabular display at the county level for Prince George County, Virginia, showing an unusual modeled 1990 acetaldehyde concentration. This observation has been marked as an overestimate after EPA's inspection of the modeled 1990 concentrations. After clicking on the "!" icon, the small explanatory message window displayed at the bottom of this figure pops up.

display) and colors have been selected to produce a pleasant visual effect. Also, icons have been incorporated that warn users of suspect numerical values. These are usually particular census tract/HAP combinations for which EPA assumes that the modeled 1990 concentrations are considerably overestimated (Figure 3). In addition to rounded data, the CEP Web site makes raw data available. Raw data are most useful for users who want to conduct additional statistical analyses.

The third navigational tool is based on so-called hierarchical clickable micromaps in the GPL environment. This approach combines LM plots and the GPL and extends their joint features to allow dynamic access to complex, geographically referenced data.

The user can mouse-click on a region (state, county) in the map and the display changes with respect to the selected region. Thus, micromaps serve as a navigational tool, but each individual LM plot is a sophisticated statistical display by itself, as described in the "Micromaps" section of this paper. The idea of using micromaps simultaneously as a statistical display and for navigational purposes, first considered for EPA's CEP Web site, may be of benefit in innumerable future applications.

In the CEP Web site, after the user selects one state in the top US micromap (or tabular display), a new micromap (or tabular display) appears—this time at the county level for the selected state. This time, the user can select a county and reach the lowest level in this hierarchy—a graphical display (or tabular display) at the census tract level. This geographic selection is an easy way to maneuver through the 60,803 census tracts, even for inexperienced users of the Web. At the higher levels (i.e., the US or state level), statistical summaries such as means, medians, minima, maxima, and quartiles are displayed for all census tracts in each region (Figure 2). At the census tract level, uncertainty bounds are displayed (Figure 3).

Creation of Micromaps

Before we can implement LM plots in the GPL environment, generalized maps that form the basis of micromaps have to be created. Previous LM plots described in the literature (3–5) use hand-created generalized maps, for example, of the United States or the countries belonging to the Organisation for Economic Cooperation and Development. The creation of such a generalized map by hand typically requires several hours of work—an option that is clearly not feasible if we are interested in generalized maps for all 50 states or all counties within the United States.

We are not aware of any existing generalized map of the 50 US states or of any of the more than 3,000 counties in the United States that satisfy our specific needs—i.e., that is available in electronic format and provides the required level of generalization. Unfortunately, regular maps are unsuitable for use in LM plots, mostly for two reasons.

First, micromaps in printed form or on a computer display typically are smaller than 2 inches by 2 inches. In this scale, a small region such as Washington, DC, would be invisible on a US map if drawn to scale. Therefore, micromaps require an exaggeration of small regions so that these regions become visible and can be color-coded.

Second, in an interactive environment such as the Web, the number of line segments determines how fast a new display is drawn and an area is filled with color. The fewer edges a map has, the faster the boundary information is passed from the Web server to the client's computer and the faster the entire graphical display is drawn.

Therefore, it is necessary to develop procedures for creating generalized maps with little user interaction. These procedures have to extract selected regions from a larger file; they also have to smooth and simplify boundaries by removing details, but keep the topological integrity of a real unit. Micromaps should not end up with holes, and neighboring regions in the original map should remain neighboring regions in the generalized map.

The creation of generalized maps for use in the CEP Web site starts with boundary files describing geographical regions and with attribute data describing the characteristics of the regions. These data, the boundary files and attribute data, are often stored in geographic information systems (GIS). In our case, we make use of ArcView (ESRI, Redlands, CA), a desktop GIS package, which is one of the most popular of GIS soft-

ware. For future use of the developed procedures for map generalization, users need to have access to ArcView and the geographical data must be available in shapefile format. Obviously, the generalized maps only have to be created once—users of the CEP Web site do not have to deal with this issue.

When creating generalized maps of the United States at the sub-state level, we start with a shapefile of the entire United States at the preferred level of geography (county, township in the New England states, census tract, or even block group). We assume that in the attribute table each record or areal unit includes a state identifier indicating to which state the areal unit belongs. A procedure written in ArcView's Avenue scripting language allows users to extract the boundary of the selected geographical level (for instance, county) by states to create a shapefile for each state. Thus, each state can be individually displayed.

Most boundary files of the United States have a relatively high level of resolution, exceeding the required resolution level for micromaps. As explained earlier, high resolution and detailed data inhibit the fast processing and display of maps on the Web. Therefore, there is a need to smooth, or simplify, the boundary by removing details but preserving the topological integrity of the areal unit.

A set of Avenue scripts based upon the Douglas-Peucker line generalization algorithm (28) has been developed to generalize boundaries to expedite the processing and display of maps on the CEP Web site. The Douglas-Peucker algorithm has been implemented in many GIS packages (including ESRI's ARC/INFO) and has been used on numerous occasions. However, the algorithm was designed to generalize linear features such as rivers and roads. It was not intended to generalize polygonal features, which is what is required for this project. The major challenge in using the Douglas-Peucker algorithm to generalize polygonal features is to maintain the topological integrity among polygons. This means that neighboring relationships among polygons have to be maintained even after the polygon boundaries have been generalized. The algorithm designed for this project can generalize polygon boundaries and, at the same time, retain the topological relationships among polygons. The detail of the algorithm is beyond the scope of this paper, but will be described and published elsewhere.

The generalization process could be performed before individual maps (by states or by counties) are extracted by the first algorithm or after each state or county file has been created. However, it is desirable to generalize maps by individual states or counties instead of the entire country because boundaries of different states have different levels of cartographic complexity. Thus, different parameter values for the generalization process have to be used to yield desirable generalization results for different states or counties. After boundaries have been generalized at the state or county level, the boundaries, in ArcView shapefile format, are converted into ASCII data in a simple format: polygons depicted by a set of points in latitude and longitude. These coordinates are later used for the micromap displays on the CEP Web site.

Implementation Issues

The CEP Web site has been designed using numerous common Web formats and styles. While the explanatory pages are mostly based on HTML files, GIF images, and PDF files for larger documents, the pages that provide access to the data are based on Java and C code, accessible through Common Gateway Interface (CGI) scripts, and automatically created HTML and JavaScript documents. Each user interaction that

results in a new display (e.g., selecting a new state or a different HAP) invokes such a C-CGI script. This script is a C program that first reads the current parameter settings (representation type, HAP, state, county) from the active Web page and then creates a new HTML/JavaScript Web page. This new page is linked once again to the same C-CGI script, but with different parameters active.

Other than two C-CGI scripts that are responsible for the top-right menu and the lower-right data display on the CEP Web site, there exists no hard-coded document that is used for the data display. Each newly visible Web page is created on the fly through the C-CGI scripts.

Developing these two C-CGI scripts required several weeks of programming time. Valuable references during the implementation process were the books by Graham (29) for HTML, Hoque for JavaScript (30), and Eckel (31) for Java. In addition, books by Weinman et al. (32,33) have been a very good source for general design issues of Web pages. The CGI used for the CEP Web site is based on code developed by Thomas Boutell and freely available on the Web at http://www.boutell.com/cgic/.

Two non-statistical Web sites have significantly influenced the design and some of the interactive features of the CEP Web site:

- http://www.usnews.com/usnews/edu/college/corank.htm; in particular, http://www.usnews.com/usnews/edu/college/rankings/natunivs/natu_a.htm, which allows users to sort university rankings according to different criteria.
- http://www.sport1.de; first click on "Fussball," then on "Bundesliga," then on "Tabelle." This is a good example of how to organize frames and update soccer standings according to different criteria (by round, home or away, etc.).

It should be noted that the CEP Web site does not use any commercial database program to access the HAP data. Because the data originate directly from a statistical package that has been used for the modeling, they have not been fed into a database program first. Instead, a three-layer tree-shaped directory structure is used; from this database, an individual data file can be directly accessed based on its state, county, and census tract federal information processing standard code.[2] Data files have been kept as small as possible—no file contains more information than the Web site needs for each newly visible Web page. This makes it unnecessary to search in files, accelerating access to the data. Also, all summary statistics at higher levels (US, county) have been precalculated to speed up access to minima, maxima, means, medians, and lower and upper quartiles.

Discussion

A first version of the CEP Web site went online in March 1998, providing general information and documents related to the project. A major remodeling of the site, providing more extensive descriptions of the air toxics data, took place in November 1998. The release of the data through the CEP Web site was scheduled to begin with interactive tables going online in December 1998. The micromap displays were expected to be posted in the following months. However, EPA ultimately decided not to make the modeled air

[2] For more information on federal information processing standard (FIPS) codes, visit http://www.epa.gov/enviro/html/codes/state.html.

toxics concentrations available through its public Web site, due to concerns that the estimates for 1990 may not be representative of current conditions. EPA has made files of the air toxics concentrations available to the public by request; data from the study have been more widely disseminated, through the Web as well as other mechanisms, by state and local environmental agencies and by other organizations. Recently, the Environmental Defense Fund added data and results based on the CEP to their http://scorecard.org Web site.

While most work on the CEP Web site was completed as originally planned, micromaps have not been fully integrated into the GPL yet. At the current stage, it seems to be advisable to revise the approach for joining the GPL with micromaps. The BLS GPL was originally developed in 1996, and a new (commercial) version of the GPL has been recently developed by an up-and-coming software company. This new GPL is currently in its alpha testing phase and it has been scheduled for release in the spring of 2000. The new GPL already contains many of the features that were intended for the BLS GPL but were never included in the old version. The inclusion of maps and micromaps based on the generalized maps developed for the CEP Web site appears to be straightforward with the new GPL. Although the BLS GPL is still a useful tool, in particular because its source code can be obtained for free from BLS, it seems to be advisable to use the new GPL for larger future applications. Several employees in federal agencies such as the federal Centers for Disease Control and Prevention and BLS have already expressed interest in investigating the use of the new GPL. Even if concerns about data timeliness mean that the CEP Web site never becomes publicly accessible, it seems to be likely that the idea of hierarchical clickable micromaps in the (new) GPL environment might be used in another federal project in the near future.

Acknowledgments

EPA funded the majority of the work behind this paper under contract Nos. 8W-0662-NAEX and 8W-1712-NTEX and cooperative agreement number CR825564-01-0. Additional federal agencies, BLS and the National Center for Health Statistics, supported some facets of this work. The article has not been subject to review by any of these agencies; it does not necessarily reflect their views, and no official endorsement should be inferred. The conclusions and opinions are solely those of the authors and are not necessarily the views of the agencies.

References

1. Carr DB. 1994. *Converting tables to plots.* Technical Report 101. Center for Computational Statistics, George Mason University, Fairfax, VA.
2. Carr DB, Olsen AR, White D. 1992. Hexagon mosaic maps for displays of univariate and bivariate geographical data. *Cartography and Geographic Information Systems* 19(4):228–36, 271.
3. Carr DB, Pierson SM. 1996. Emphasizing statistical summaries and showing spatial context with micromaps. *Statistical Computing and Statistical Graphics Newsletter* 7(3):16–23.
4. Carr DB, Olsen AR, Courbois JP, Pierson SM, Carr DA. 1998. Linked micromap plots: Named and described. *Statistical Computing and Statistical Graphics Newsletter* 9(1):24–32.

5. Carr DB, Olsen AR, Pierson SM, Courbois JP. Using linked micromap plots to characterize Omernik ecoregions. *Data Mining and Knowledge Discovery.* To appear.
6. Carr DB, Valliant R, Rope D. 1996. Plot interpretation and information webs: A time-series example from the Bureau of Labor Statistics. *Statistical Computing and Statistical Graphics Newsletter* 7(2):19–26.
7. Symanzik J, Axelrad DA, Carr DB, Wang J, Wong D, Woodruff TJ. 1999. *HAPs, micromaps and GPL—Visualization of geographically referenced statistical summaries on the World Wide Web.* American Congress on Surveying and Mapping, Annual Proceedings of the ACSM-WFPS-PLSO-LSAW 1999 Conference. March 13–17, 1999. CD-ROM.
8. US Environmental Protection Agency. 1998. *National air quality and emissions trends report, 1997.* Research Triangle Park, NC: US Environmental Protection Agency. EPA 454/R-98-016.
9. Ministry of Public Health. 1954. *Mortality and morbidity during the London fog of December 1952.* Reports on Public Health and Medical Subjects, No. 95. London, England: Her Majesty's Stationery Office.
10. Schrenk H, Heimann J, Clayton G, Gafafer W, Wexler H. 1949. *Air pollution in Donora,* PA. Public Health Service Bulletin No. 36. Washington, DC: Public Health Service.
11. US Environmental Protection Agency. 1994. *Technical background document to support rulemaking pursuant to Clean Air Act Section 112(g): Ranking of pollutants with respect to human health.* Research Triangle Park, NC: US Environmental Protection Agency. EPA-450/3-92-010.
12. Glickman T, Hersh R. 1995. *Evaluating environmental equity: The impacts of industrial hazards on selected social groups in Allegheny County, Pennsylvania.* Discussion Paper 95-13. Washington, DC: Resources for the Future.
13. Perlin SA, Setzer RW, Creason J, Sexton K. 1995. Distribution of industrial air emissions by income and race in the United States: An approach using the Toxic Release Inventory. *Environmental Science and Technology* 25:69–80.
14. Cote I, Vandenberg J. 1994. Overview of health effects and risk-assessment issues associated with air pollution. In: *The vulnerable brain and environmental risks. Vol. 3: Toxins in air and water.* Ed. R Isaacson, K Jensen. New York: Plenum Press.
15. Cupitt L, Cote I, Lewtas J, Lahre T, Jones J. 1995. *EPA's Urban Area Source Research Program: A status report on preliminary research.* Washington, DC: US Environmental Protection Agency. 600/R-95/027.
16. Hassett-Sipple B, Cote I, Vandenberg J. 1991. Toxic air pollutants and non-cancer health risks—United States and a Midwestern urban county. *Morbidity and Mortality Weekly Report* 40:278–80.
17. Office of Air Quality Planning and Standards. 1990. *Cancer risk from outdoor exposure to air toxics.* Research Triangle Park, NC: US Environmental Protection Agency. EPA-450/1-90/004a.
18. Kelly T, Mukund R, Spicer C, Polack A. 1994. Concentrations and transformations of hazardous air pollutants. *Environmental Science and Technology* 28:378A–87A.
19. Woodruff T, Axelrad D, Caldwell J, Morello-Frosch R, Rosenbaum A. 1998. Public health implications of 1990 air toxics concentrations across the United States. *Environmental Health Perspectives* 106:245–51.
20. Rosenbaum A, Ligocki M, Wei Y. 1999. *Modeling cumulative outdoor concentrations of hazardous air pollutants.* Revised Final Report. San Rafael, CA: Systems Applications International, Inc. http://www.epa.gov/CumulativeExposure.
21. Caldwell J, Woodruff T, Morello-Frosch R, Axelrad D. 1998. Application of hazard identification information for pollutants modeled in EPA's Cumulative Exposure Project. *Toxicology and Industrial Health* 14:429–54.

22. Anderson G. 1983. *Human exposure to atmospheric concentrations of selected chemicals.* Vol. 1. US Environmental Protection Agency. NTIS PB84-102540.
23. US Environmental Protection Agency. 1991. *Toxics Release Inventory 1987–1990.* Electronic Version on CD-ROM. Washington, DC: US Environmental Protection Agency.
24. US Environmental Protection Agency. 1993. *Regional interim emission inventories (1987–1991), volume 1: Development methodologies.* Research Triangle Park, NC: US Environmental Protection Agency. EPA-454/R-93-021a.
25. EH Pechan and Associates. 1994. *Emissions inventory for the National Particulate Matter Study.* EPA Contract 68-D3005. Springfield, VA: EH Pechan and Associates.
26. Cleveland WS. 1993. *Visualizing data.* Summit, NJ: Hobart Press.
27. Cleveland WS. 1994. *The elements of graphing data.* Summit, NJ: Hobart Press.
28. Douglas DH, Peucker TK. 1973. Algorithms for reduction of number of points required to represent a digitized line or its character. *The Canadian Cartographer* 10:112–22.
29. Graham IS. 1998. *HTML 4.0 sourcebook.* 4th Ed. New York, NY: John Wiley and Sons.
30. Hoque R. 1997. *Practical JavaScript programming.* New York, NY: M&T Books.
31. Eckel B. 1998. *Thinking in Java.* Upper Saddle River, NJ: Prentice Hall.
32. Weinman L. 1996. *Designing Web Graphics.2.* Indianapolis, IN: New Riders.
33. Weinman L, Lentz JW. 1998. *Deconstructing Web Graphics.2.* Indianapolis, IN: New Riders.